

A kauzális kohézió vizsgálata az Intex számítógépes eszközzel

Mészáros Ágnes¹

Papp Orsolya²

¹SE Pszichiátriai és Pszichoterápiás Klinika
H-1083 Budapest, Balassa u. 6.
meszaros@psych.sote.hu

²PTE BTK Pszichológia Doktori Iskola
H-7624 Pécs, Ifjúság útja 6.
papporsi@lycos.com

1. Bevezetés

A tanulmány oksági kapcsolatok természetes nyelvi szövegekben történő automatikus azonosításának egy kísérletét mutatja be. Az MTA Pszichológiai Kutatóintézet 2004 tavaszán felvett élettörténeti szövegkorpuszát⁷⁸ az Intex számítógépes eszköz [9] segítségével elemeztük, annak a szemléleti és módszertani hagyománynak a talaján, mely az elbeszélő bizonyos személyiségbeli tulajdonságaira az elbeszélésbeli események leírásának szó- és mondat szinten megjelenő nyelvi-formai jegyeiből következtet [12, 3].

Schank és Abelson [8] óta a kognitív pszichológiában a világról és önmagunkról alkotott fogalomrendszerünk elsődleges szervező elveként tételezett oksági kapcsolatok vizsgálata és formalizált leírás kísérlete természetes módon illeszkedik bele az identitást az önmagunkról alkotott történetek, céljaink szerint változó összességeként leíró narratív pszichológiai kutatás azon törekvésebe, hogy az önéletrajzi elbeszélés egyes dimenzióit (például kauzális szerveződését) automatizált eszközök segítségével ragadja meg. A pszichológiai kérdésfeltevés tehát a következő elemekből tevődik össze: az okság típusainak konceptualizálása és pszichológiai mérőeszközeinek feltérképezése az egyik oldalon, a másikon pedig a narratívumok összefüggésrendszerét megteremtő oksági jegyek nyelvi operacionalizálása és automatikus felismerésükre tartalomelemző szoftver fejlesztése [1]. Tanulmányunkban a következő pontban az oksági kategóriákat Trabasso és van den Broek [11] modellje alapján tárgyaljuk, ezután a nyelvi markerek azonosítására és automatikus felismerésére teszünk javaslatot.

⁷⁸ A vizsgálatban résztvevő személyek hat személyes narratívumot idéztek fel: első emléküket, teljesítményt, veszteséget, félelmet, egy fontos személyhez kötődő jó és rossz kapcsolati élményt.

1.1 A kauzalitás egy modellje és mérési lehetőségei

Az oksági kapcsolatok kontextusfüggő, logikai-szemantikai kritériumok alapján végzett vizsgálatára szoros történetvezetésű, tanulásra kiélezett mesékben Trabasso és mtsai tanulmányai [12, 13] szolgálhatnak mintául. A narratívumok egyes eseményeinek fontosságát és felidézhetőségét az esemény manuálisan azonosított közvetlen oksági kapcsolatainak számával, valamint egy oksági láncon való elhelyezkedésével hozták összefüggésbe. Elemzési módszerük különböző korpuszokon történő alkalmazása azonban arra a következtetésre vezette őket, hogy nem lehet figyelmen kívül hagyni a vizsgált narratívum makrostruktúráját, a történetnyelvtanok által azonosított rekurzív szerkezeti elemek sajátos felépítési sorrendjét. Rekurzív tranzíciós hálózat modelljükben [11] ennek folyományaként a narratívumokat véges eseménykategoróriák (szetting, cél, elérési kísérlet, reakció, kimenetel) és oksági tipológiába rendezett viszonyok láncolataként írják le. Mivel korábban a narratív koherencia vizsgálatakor [6] amellet érveltünk, hogy csupán a történetstruktúra speciális szerkezeti elemeinek azonosítási képessége nyújthatna biztos kiindulópontot a narratív pszichológiai kérdések megválaszolásában, az ennek részét képező kauzális koherencia elemzésekor is egy olyan modellt preferálunk, mely a szó- és mondat szintű elemek mellett a narratívum makrostruktúráját is figyelembe veszi. Trabasso és van den Broek modellje ezek alapján megfelelő elméleti keretként szolgál különböző oksági viszonyok⁷⁹ felismeréséhez, nem nyújt azonban kidolgozott támpontokat szövegek automatizált, számítógépes elemzéséhez.

A Memphisi Egyetem Pszichológiai Kutatócsoportja által angol nyelvre kifejlesztett COH-METRIX szövegelemző program felépítése és általános működési elvei [2] ebből a szempontból úttörő szerepet töltenek be. A szoftver intencionalitás moduljától elkülönített kauzalitás modul szószinten vizsgálja az egyes (nem feltétlenül narratív) szövegek kohézióját egy olyan aránypárral jellemezve az elemzett szövegeket, melyben a számláló az oksági viszonyokat jelző partikulák (specifikus kötőszavak, például *so that*, *because*, valamint az oksági viszony jelenlétére általánosságban utaló igék és segédigék, például *cause*, *make*) számát tartalmazza, a nevező pedig a WordNet szemantikai hálóban oksági jeggyel jelölt igék (például *kill*) számát.

Magyar nyelvre fejlesztett, természetes nyelvű önéletrajzi narratívumokat elemző programunk távlati célja éppen az automatizált szóalapú és a logikai-szemantikai kiindulású, manuális elemzések közötti eltérések kezelése. Előadásunkban ennek első lépéseit mutatjuk be, az Intex számítógépes eszközzel kifejlesztett, Trabasso és van den Broek [11] egyes oksági típusai mentén összegyűjtött szavak és kifejezések bizonyos szintaktikai és szemantikai tulajdonságait is figyelembe vevő gráfok működésén keresztül.

⁷⁹ Az általunk végzett elemzéshez az eredeti tipológiából a Motivációs (cél kifejező szavak jelenléte), a Pszichológiai (belső állapotváltozást, reakciót leíró szavak, kifejezések), valamint a Fizikai/Eredmény típusú okságot (két cselekvést vagy egy cselekvést és egy állapotváltozást kifejező ige egymáshoz közeli jelenléte) használjuk fel. Az eredeti modell részletes leírását lásd Trabasso és van den Broek [11], a változtatások indoklását pedig a tavalyi kötetben [7].

1.2 A kauzalitás nyelvi megjelenése

Az oksági kapcsolat alapvetően proposíciók *közötti* viszonyra vonatkozik, ami a történetekben két esemény közötti relációt feltételez. Trabasso és van den Broek modelljében ez tagmondatpárok vizsgálatában jelentkezik olyan módon, hogy az egyik esemény (esemény1) az egyik, míg a másik (esemény2) a következő tagmondatban található. További megszorításokat is alkalmaz a nyelvész az igei oksági kapcsolat jellemzésére [4]:

esemény1 *oka* esemény2-nek, ha

1. esemény1 implikálja esemény2-t,
2. esemény2 időben nem előzi meg esemény1-et,
3. és esemény1 jelenléte szükséges feltétele esemény2 bekövetkezésének (kontrafaktuális feltétel).⁸⁰

Elemzésünk kettős irányú: a szövegtörzs manuális vizsgálatakor, amely a számítógépes alkalmazás validálása egyben, a fenti három feltétel teljesülésekor azonosítunk oksági kapcsolatot. Azonban modulunk jelenlegi felépítése nem párok azonosítására alkalmas, csupán egyes szavakra vagy több szóból álló kifejezésekre kereshetünk vele. A logikai kritériumok érvényesülése pedig ige-párokból álló lexikonok fejlesztésén keresztül valósulhat meg.

1.3 Az elemzés eszköze

Az elemzést az Intex számítógépes eszközzel végezzük [9], melyet természetes nyelvek formalizált leírására fejlesztettek ki. Az eszköz nagyméretű korpuszok valószínű feldolgozását teszi elérhetővé: a gyors elemzés lehetőségét az nyújtja, hogy mind a bemeneti szöveget, mind a nyelvtanokat tömörített formában tárolja.

A programot lexikalista nyelvelemzésre fejlesztették ki, melynek alapjai az elektromos szótárak, és véges állapotú gráfokban ábrázolt nyelvtani szabályok. A szótárakban a szavak morfoszintaktikai és szemantikai jegyei egy szinten vannak kódolva, így a gráfokban mindezen információra egyidejűleg lehet hivatkozni. Az eszközt adottságai különösen alkalmassá teszik tartalomelemzési feladatokra, hiszen lehetőség van előre összeállított lexikonok elemeinek – akár szemantikai jegyek alapján történő – együttjárási vizsgálatára.

2. Az okság–modul felépítése

Az okságra vonatkozó elméleti modell kategóriáinak megfelelően három gráfot szerkesztettünk az egyes kauzalitás típusok, a történetben cél jelenlétét jelző Motivációs,

⁸⁰ Trabasso és van den Broek az oksági típusok manuális azonosításában használják ezeket a logikai feltételeket, elsősorban a kontrafaktuális tesztre támaszkodnak. Ezzel szemben például a WordNet szemantikai rendszerben a harmadik kritérium teljesülésétől bizonyos ige-párok esetében eltekintenek (például *give–have*) [5].

belső állapotra vagy reakcióra utaló Pszichológiai, illetve két esemény közötti Fizikai viszony felismerésére. Jelen tanulmányban az utóbbi két gráfot mutatjuk be részletesen, az ágens intencionális hozzáállását jelző igei szerkezeteket és főneveket azonosító Motivációs okságot korábban [7] vizsgáltuk. Hangsúlyoznunk kell, hogy a modul kialakításának jelenlegi stádiumában kizárólag az adott korpuszban előforduló szavak és kifejezések feldolgozására törekszünk.

2.1 A pszichológiai okság azonosítása

A pszichológiai okság azonosítását az emocionális állapotokat leíró szavak felől közelítettük., melyek lehetnek önmagukban álló szemantikailag specifikus főnevek (például *fájdalom*, *katasztrófa*), melléknevek (például *szörnyű*, *boldog*), igék (*csalódik*, *szorong*), illetve az érzés/érez szó előtt álló melléknevek vagy határozószók.

Az alábbi példák ezen gráf találatainak természetes nyelvi környezetben való előfordulásait mutatják:

„Tehát például amikor a legjobb barátom 2002. december 21.-én öngyilkos lett.{S} Ez egy elég súlyos veszteség volt.”

„Amszterdamban is előfordult, hogy a lánynak beszólt egy fiú az utcán, egy tök ismeretlen férfi, és a lány elkezdett futni utána, a srác az pedig legyintett, hogy 'úristen, most mit csináljak, hát most akkor ezzel mit kezdjek, fussak utána, hát csináljon, amit akar', de közben folyamatosan szenvedett és dühös volt.”

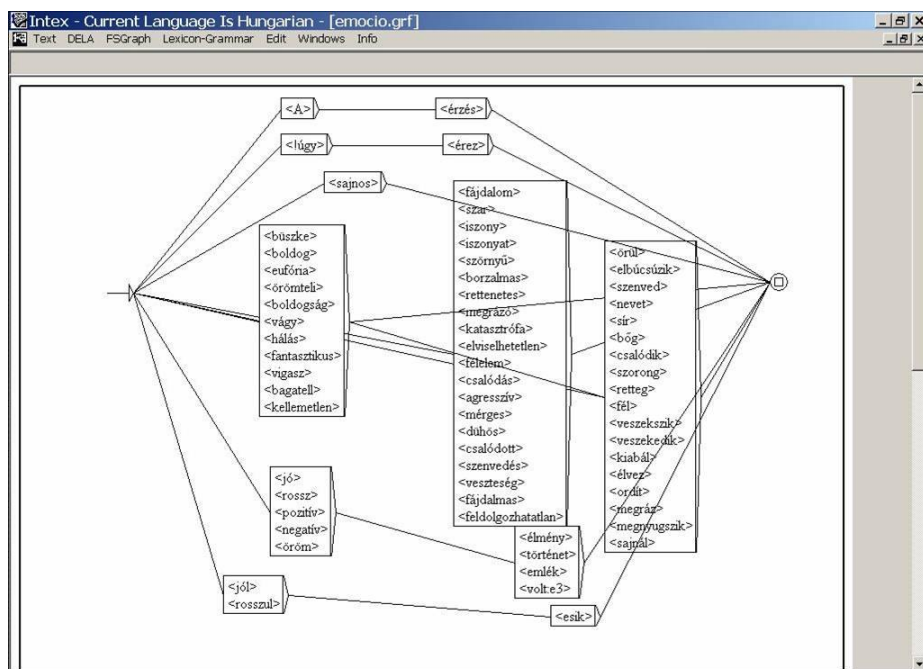
„És ez ilyen nagyon jó érzéssel töltött el, hogy bebizonyítottam a családomnak, hogy igen is, ha nagyon akarom, akkor meg tudom csinálni.”

Ebbe a kategóriába soroltuk a történet egészére vonatkozó értékelő reflexiókat is, melyek érzelmi viszonyulást implikálnak.

„...és akkor még utána az orvos is tette a megjegyzést, és akkor ugye az volt a probléma igazából, nem maga, hát jó, maga a cselekedet, tehát maga az abortusz volt a baj, de az, hogy végül is nem sok időnk volt rá, eldönteni, és tulajdonképpen azt beszéltük, hogy ha nem most lett volna, hanem később, akkor akár igen. És ezért ez így rossz volt.”

*„És én óvadás voltam és valami úton-módon nem szerettem az óvodában aludni. Állandóan fönt voltam, magyarul rossz voltam.”

Az utolsó példa annak szükségességét mutatja, hogy a történet egészére vonatkozó értékelő reflexióknál a létigének csak E/3 alakját vehetjük figyelembe.



1. ábra. Érzelmi állapotokat azonosító gráf

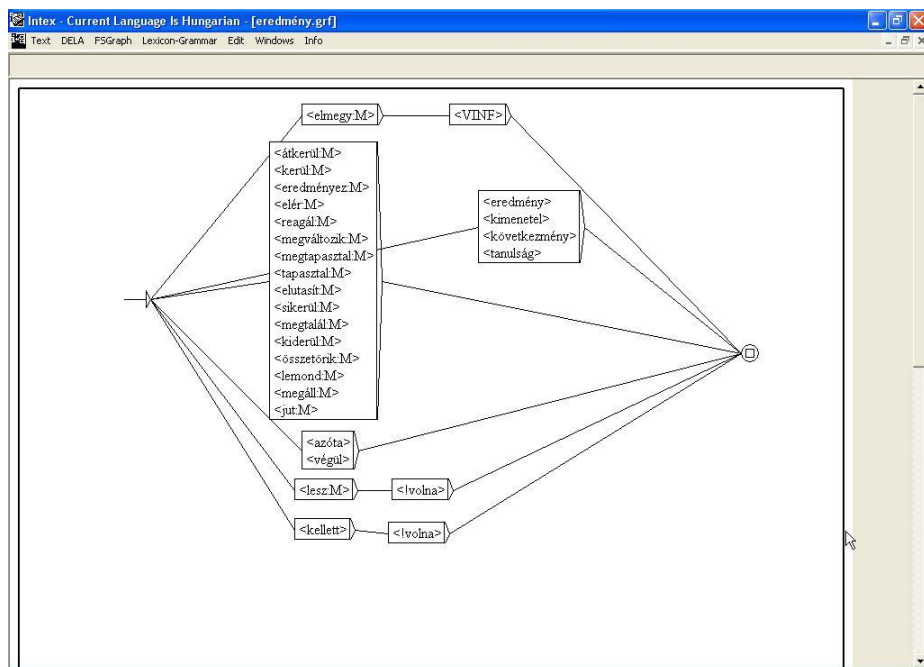
2.2 A fizikai okság azonosítása

Ennél az oksági kategóriánál szerencsésebbnek tartjuk a Schank és Abelson [8] által használt Eredmény okság elnevezést, mivel sokkal inkább utal az események közötti kontrafaktuális viszony természetére, mely nem csupán fizikai lehet.

Az Eredmény típusú oksági viszony azonosítását szolgáló gráf szavait a vizsgált korpuszon a kontrafaktuális teszt alapján végzett manuális elemzés szerint választottuk ki, és csupán azokat hagytuk benne a jelenlegi szövegtárban, melyek legalább 70 %-ban helyes találatokhoz vezettek. A keresett szavak túlnyomó részben igék, melyek múlt idejű, nem feltételes módú alakban fordultak elő a korpuszban (például *kiderült*, *sikerült*); emellett szemantikailag specifikus főnevek (például *eredmény*, *tanulság*), valamint néhány funkciószó (például *azóta*, *kellett*) jelölte az eredmény okságot.

„...nem vettek fel elsőre az orvosira, így elmentem dolgozni nővérként.”

„Talán az utóbbi időkből az egyik legrosszabb történet a fiammal kapcsolatban, amikor annak ellenére, hogy nem volt neki engedélyezve, az autót kivitték, elmentek furikázni, főtengelyesre hajtották, jó pár ezer forintba került, ez pár évvel ezelőtt fordult elő.”



2. ábra. Eredmény típusú oksági viszonyt azonosító gráf

Az 'element'+főnévi igenév-szerkezet előfordulási jellegzetessége, hogy az eredmény típusú oksági viszony részeként egyes esetekben a kiinduló eseményt (2. példa), máskor a következményt (1. példa) jelzi. Az alábbi példa azt mutatja, hogy a rossz találatok strukturálisan a történet bevezető részében (setting) fordultak elő.

*„Volt egy kirándulás, ami nagyon negatív eredménnyel zárult, ugyanis apáék elementek vadászni, és a húgommal, ez télen volt, ez Szabolcs megyében történt, és ott nagyon ritka a domb, és volt egy nagy domb, és vittük a szánkót, és a húgommal szánkóztunk.”

A 'kellett' segédige szövegbeli megjelenése –performativitása miatt– 85%-ban Eredmény típusú oksági viszonyt azonosított. A találatokat részletesen megvizsgálva azt találtuk, hogy a főnévi igenéves szerkezettel olyan eredményekről számolnak be a beszélők, amelyek külső hatásra következtek be.

„De például ez az ismerősöm kiállt mellettem abban, hogy ne higgyek másoknak, saját magam tapasztalataim alapján vonjak le következtetéseket és végülis így is tettem, és tényleg abszolút nem is kellett csalódnom.”

Az idézett példa az ún. kettős kódolás kérdését is felveti, amennyiben az eredményként értékelhető reakció emocionális válaszként is felfogható.

A 'sikertül' múlt idejű alakja 82%-ban adott helyes találatot a vizsgált korpuszban az eredmény típusú okságra, téves azonosítás akkor fordult elő, amikor kognitív ige állt a közvetlenül előtte levő tagmondatban.

„Hát mondjuk a legutóbbi, az a Katinak az esküvője, azt szerveztem, és azok jól sikerültek.”

*„És elkönyveltem magamban, hogy akkor ez most nem sikerült, aztán kiderült, hogy mégis.”

Az *azóta* kötőszóval kapott találatok részletes elemzése szintén indokolt az eredmény típusú oksági viszonyok azonosításakor, hiszen funkciója szerint az egyik tagmondatban leírt, múltbeli eseményt köti össze egy másik tagmondatral, amelyben a jelenre gyakorolt hatásról van szó. Értelemszerűen oksági kapcsolat áll fenn a két esemény között, a múltbeli esemény az oka annak, ami a jelenben zajlik.

Korpuszunkban 31 találatot kaptunk. Az esetek túlnyomó részében (24) helyes, tehát eredmény típusú oksági viszonyt jelöl az *azóta* kötőszóval összekapcsolt két tagmondat. A jó találatokban az ige jelen idejű az eredmény oldalon, ami jól mutatja a jelenre gyakorolt hatást:

„...és megismerkedtem egy olyan fiúval, aki viszont árva volt, tehát ő is egyedül volt, és tizenkilenc éves koromban, és azóta tart a szerelem, az együttlét, mert ő a férjem,...

A téves találatokban a kötőszót követő ige gyakrabban múlt idejű.

„És talán ilyen erős félelmet azóta sem éreztem.”

Közös jellegzetességük ezen felül, hogy nem két esemény között teremt kapcsolatot a kötőszó, hanem a múltbeli esemény jelentőségét hangsúlyozza a beszélő:

*„Szóval ekkor féltem életemben egyszer, először és talán azóta is utoljára.”

2.3 Kitekintés

Az Eredmény típusú okság azonosításában mindenképpen további cél egy olyan korpuszkiindulású igepár–lista felállítása, mely nem általános oksági szemantikájú, hanem tartalmas szavak mentén (például *lelőtte*→*meghalt*) eredményez találatokat. A pszichológiai irányú kérdésfeltevések szempontjából pedig a következő lépés a mentális folyamatra utaló 'kognitív' igék és kifejezések elhelyezése az események által kiváltott belső reakciók kategóriáján belül.

Bibliográfia

1. Ehmann, B., Kis, B., Naszódi, M., László, J.: A szubjektív időélmény tartalomelemzéses vizsgálata. *Pszichológia*, Vol. 25 (2), 2005, 133-142
2. Graesser, A.C., McNamara, D.S., Louwerse, M.M., Cai, Z.: Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments and Computers* (2004)
3. László, J., Ehmann, B., Péley, B., Pólya, T.: A narratív pszichológiai tartalomelemzés: elméleti alapvetés és első eredmények, *Pszichológia*, Vol. 20, 2000, 367-390
4. Kiefer F. : Jelentélmélet, Corvina, Budapest (2000)
5. Kuti, J., Vajda, P., Varasdi, K.: Javaslat a magyar igei WordNet kialakítására. III. Magyar Számítógépes Nyelvészeti Konferencia konferenciakötete, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged (2005)
6. Papp, O.: Narratív koherencia-elemző program helye a pszichológiai kutatásban. II. Magyar Számítógépes nyelvészeti konferencia, Konferenciakötet, Kiadó: Juhász Nyomda, Szeged (2004)
7. Papp, O., Mészáros, Á.: Oksági viszonyok azonosítása önéletrajzi narratívumokban. III. Magyar Számítógépes Nyelvészeti Konferencia konferenciakötete, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 2005
8. Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals and understanding: an Inquiry into human Knowledge Structures. Erlbaum, Hillsdale (1977)
9. Silberstein, M.: Intex Manual. Internetes elérhetőség: www.nyu.edu/pages/linguistics/intex/
10. Trabasso, T., van den Broek, P.: Causal Networks versus Goal Hierarchies in Summerizing Text. *Discourse Processes*, Vol. 9. Alex Publishing Corporation, New Jersey (1986) 1-15
11. Trabasso, T., van den Broek, P., Suh, S.Y.: Logical Necessity and Transitivity of Causal Relations in Stories. *Discourse Processes*, Vol. 12. Alex Publishing Corporation, New Jersey (1989) 1-25
12. Trabasso, T., Sperry, L. L.: Causal Relatedness and Importance of Story Events. *Journal of Memory and Language*, Vol. 24 (5), 1985, 595-611.
13. Trabasso, T., van den Broek, P.: Causal Thinking and the Representation of Narrative Events, *Journal of Memory and Language*, Vol. 24 (5), 1985, 612-630.